Protein Science (1997), 6:2166–2179. Cambridge University Press. Printed in the USA.

Ligand binding to proteins: The binding landscape model

DAVID W. MILLER[1] AND KEN A. DILL[1,2]

[1]Graduate Group in Biophysics, University of California at San Francisco, San Francisco, California 94143-1204
[2]Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-1204

(RECEIVED March 18, 1997; ACCEPTED May 28, 1997)

Abstract

Models of ligand binding are often based on four assumptions: (1) steric fit: that binding is determined mainly by shape complementarity; (2) native binding: that ligands mainly bind to native states; (3) locality: that ligands perturb protein structures mainly at the binding site; and (4) continuity: that small changes in ligand or protein structure lead to small changes in binding affinity. Using a generalization of the 2D HP lattice model, we study ligand binding and explore these assumptions. We first validate the model by showing that it reproduces typical binding behaviors. We observe ligand-induced denaturation, ANS and heme-like binding, and "lock-and-key" and "induced-fit" specific binding behaviors characterized by Michaelis–Menten or more cooperative types of binding isotherms. We then explore cases where the model predicts violations of the standard assumptions. For example, very different binding modes can result from two ligands of identical shape. Ligands can sometimes bind highly denatured states more tightly than native states and yet have Michaelis–Menten isotherms. Even low-population binding to denatured states can cause changes in global stability, hydrogen-exchange rates, and thermal *B*-factors, contrary to expectations, but in agreement with experiments. We conclude that ligand binding, similar to protein folding, may be better described in terms of energy landscapes than in terms of simpler mass-action models.

Keywords: global stability; HP lattice model; hydrogen exchange; ligand binding; protein flexibility

Standard assumptions in binding models

Ligand binding is important for protein function. A quantitative understanding of many biological binding processes has been gained through binding polynomial models (Wyman & Gill, 1990; Ackers et al., 1992; Di Cera, 1995), empirical equations that relate the fractional occupation of binding sites to free ligand concentration. This approach underlies Michaelis–Menten kinetics, Hill and Scatchard plots, and cooperativity and allostery as embodied in the MWC (Monod et al., 1965) and KNF (Koshland et al., 1966) models, for example. Despite the great power and widespread usage of binding polynomial models in biochemistry, they are incomplete and phenomenological. For a given protein and ligand, binding polynomial models do not tell us where or how tightly the ligand will bind, whether or not binding will induce a conformational change, or whether the ligand will denature the protein, for example. Binding polynomial models begin by *assuming* some mass-action scheme for the binding process, and the binding and cooperativity constants are then determined by curve-fitting to experimental data. Finding the right binding model for a given ligand–protein system is a matter of trial and error.

A more complete binding model would *predict* the binding sites, the binding constants and cooperativity, and the perturbations of the protein, based on knowledge of the ligand structure and the amino acid sequence of the protein. Such models are far beyond the current scope of computational biochemistry. Nevertheless, we take a step in that direction by using a simplified statistical mechanical model of protein–ligand interactions, for which we can exactly enumerate all possible protein conformations and binding modes. Our aim here is to describe such a model and its predictions for binding.

First, we show that the model leads to many of the familiar types of protein–ligand binding, including "lock and key" and "induced fit" specific binding, ANS binding to molten globules, and ligand-induced denaturation, among others. We then explore some interesting and unconventional behaviors predicted by the model, many of which are not interpreted readily using simpler mass-action models. In particular, we address four premises in which the current paradigm of ligand binding is heavily rooted: (1) Binding is largely *shape-determined*, as embodied in the terms "lock-and-key" and "induced fit." (2) Ligands bind principally to the *native states* of proteins with little or no interaction with the *unfolded* states. Here is a typical description (Creighton, 1993): "A general consequence of ligand binding is that the protein is stabilized against unfolding and is less flexible . . . [This is] a consequence of the ligand binding more tightly to the fully folded conformation (N) than to the fully unfolded state (U) and any distorted or partially unfolded forms that result from flexibility of the structure." (3) Binding is highly *localized*. The main perturbations of the protein structure are assumed to be near the binding site. (4) Small

Reprint requests to: Ken A. Dill, Department of Pharmaceutical Chemistry, Box 1204, University of California at San Francisco, San Francisco, California 94143-1204; e-mail: dill@maxwell.ucsf.edu.

changes in the structure of a ligand or protein lead to only small changes in the bound complex.

These premises are usually supported by X-ray and NMR structures of unbound and complexed proteins, and by the successes of structure-based drug design methods (Kuntz, 1992; Shoichet et al., 1993; Bohacek & McMartin, 1994; Strynadka et al., 1996). Nevertheless, a few recent experimental results, particularly from hydrogen exchange, are puzzling when interpreted using these premises. Here we develop a model that, unlike binding polynomial models, aims to connect structure to thermodynamics. Although the physical model is simple, the statistical mechanics are rigorous so we can test such premises, rather than assume them. For reasons that will become clear below, we call ours the Binding Landscape Model, to contrast it to those based on the premises above, such as the Lock-and-Key and Induced Fit models. The groundwork for connecting ligand binding to energy landscapes has been described in theoretical and experimental work on small molecule binding to globins by Frauenfelder, Wolynes, and others (Frauenfelder et al., 1991).

Modeling proteins using the HP lattice model

We model proteins using the two-dimensional HP lattice model (Lau & Dill, 1989, 1990; Chan & Dill, 1991; Dill et al., 1995). A protein is represented as a sequence of H (hydrophobic) and P (other) monomers on a two-dimensional lattice. Lattice sites may be either empty or filled by a single monomer, and empty lattice sites are assumed to contain a solvent molecule. Each HH contact, formed when two nonsequential H monomers occupy adjacent lattice sites, is favored by a free energy ϵ ($\epsilon < 0$), which is meant to capture the importance of hydrophobic interactions in protein collapse and global stability (Dill, 1990). Hence, the free energy of a conformation is $h\epsilon$, where h is the number of HH contacts. The magnitude of ϵ determines the stability imparted by external conditions: large and negative ϵ reflects conditions that are more stabilizing, such as lower temperature or lower denaturant concentrations. Conformational entropy, the driving force for unfolding, enters the model through the exhaustive enumeration of all the possible chain configurations (see below).

The disadvantages of the model are clear: atomic resolution is lost; conformations are restricted to a lattice; it is in two dimensions; the energy function is simplified; and chains are unrealistically short. Yet, despite these disadvantages, the model has been found useful for modeling protein properties (Chan & Dill, 1989, 1990, 1994; Lau & Dill, 1989, 1990; Shortle et al., 1991; Dill et al., 1995; Miller & Dill, 1995) because it shows several protein-like features, including cooperative collapse, native structures having a nonpolar core and definable secondary structures, multi-stage folding kinetics, and molten globule states. Most importantly, we believe the model captures the main physics of protein folding—the hydrophobic interactions, conformational freedom of the chain, and the steric restrictions imposed by excluded volume.

We study HP sequences having 16 monomers. For any 16-mer chain, there are exactly 802,075 possible conformations that can be configured on a two-dimensional square lattice. These conformations are generated by computer, and each is weighted by a Boltzmann factor according to the number of HH contacts made. Figure 1A shows an energy diagram for a sample HP sequence (called "sequence A"). The native structure (ground state) is the conformation with the largest possible number of HH contacts, and thus the lowest free energy at low temperatures. We study only non-degenerate sequences, i.e., those having a single native conformation, because we believe they best represent biological proteins, which fold to unique structures. All higher-energy conformations comprise the non-native, or "denatured" states, and are grouped by energy into "first-excited" states, "second-excited" states, etc., corresponding to successively fewer HH contacts. For any HP sequence, there are far more open, high-energy conformations than compact, low-energy conformations (see Fig. 1B).

Modeling the ligand and its interactions with the protein

We model ligands as single, monomer-sized beads (Fig. 1C). A protein–ligand contact occurs when the ligand occupies a lattice site adjacent to a chain monomer. Here we consider only nonpolar ligands: a contact between a ligand and an H monomer (LH contact) is favored by a free energy $b\epsilon$, where ϵ is the HH contact energy, and b is a positive constant ($0 < b < 1$). In order to have the simplest possible model of binding, we assume the interaction energy is zero between a ligand and a P monomer, and zero between ligands. The total contact energy, E_s, for any protein–ligand configuration ("ligation state"), s, is therefore

$$E_s = h\epsilon + mb\epsilon, \tag{1}$$

where m is the total number of LH contacts. The ligation state in Figure 1C has three HH contacts and six LH contacts, so the total energy is $3\epsilon + 6b\epsilon$.

Below we show how average properties of protein–ligand complexes are derived through exact enumeration of all the possible protein–ligand configurations. First, we present the general statistical mechanical theory, which is independent of the lattice or any other specific model. Then, we introduce the HP lattice model to relate the binding thermodynamics to the corresponding structures. A more detailed derivation of the theory is given in the Appendix.

Statistical mechanics of ligand binding

The probability, P_s, of any protein–ligand ligation state, s, is given by the grand canonical distribution function:

$$P_s = \frac{e^{-E_s/kT} e^{\mu N_s/kT}}{\sum_{s=1}^{\Gamma} e^{-E_s/kT} e^{\mu N_s/kT}}, \tag{2}$$

where T is absolute temperature, μ is the ligand chemical potential, E_s is the total energy of the ligation state (i.e., due to intrachain contacts plus ligand contacts), N_s is the number of ligands bound, k is Boltzmann's constant, and Γ is the total number of ligation states available to the protein–ligand complex.

Equation 2 gives the probability of a specific ligation state as a function of temperature and ligand concentration (related to the ligand's chemical potential; see Appendix). In broad terms, this equation predicts the following behaviors. At low temperature and low ligand concentration (large negative μ), proteins are folded and have few ligands bound. Increasing the ligand concentration lowers the unfavorable translational entropy of binding, and more ligands bind. Increasing the temperature denatures the protein. It is the interplay of these behaviors, for different ligands and different proteins, that we explore more fully below.

Equation 2 is general and model-independent. It permits the calculation of various average protein properties (see Appendix),

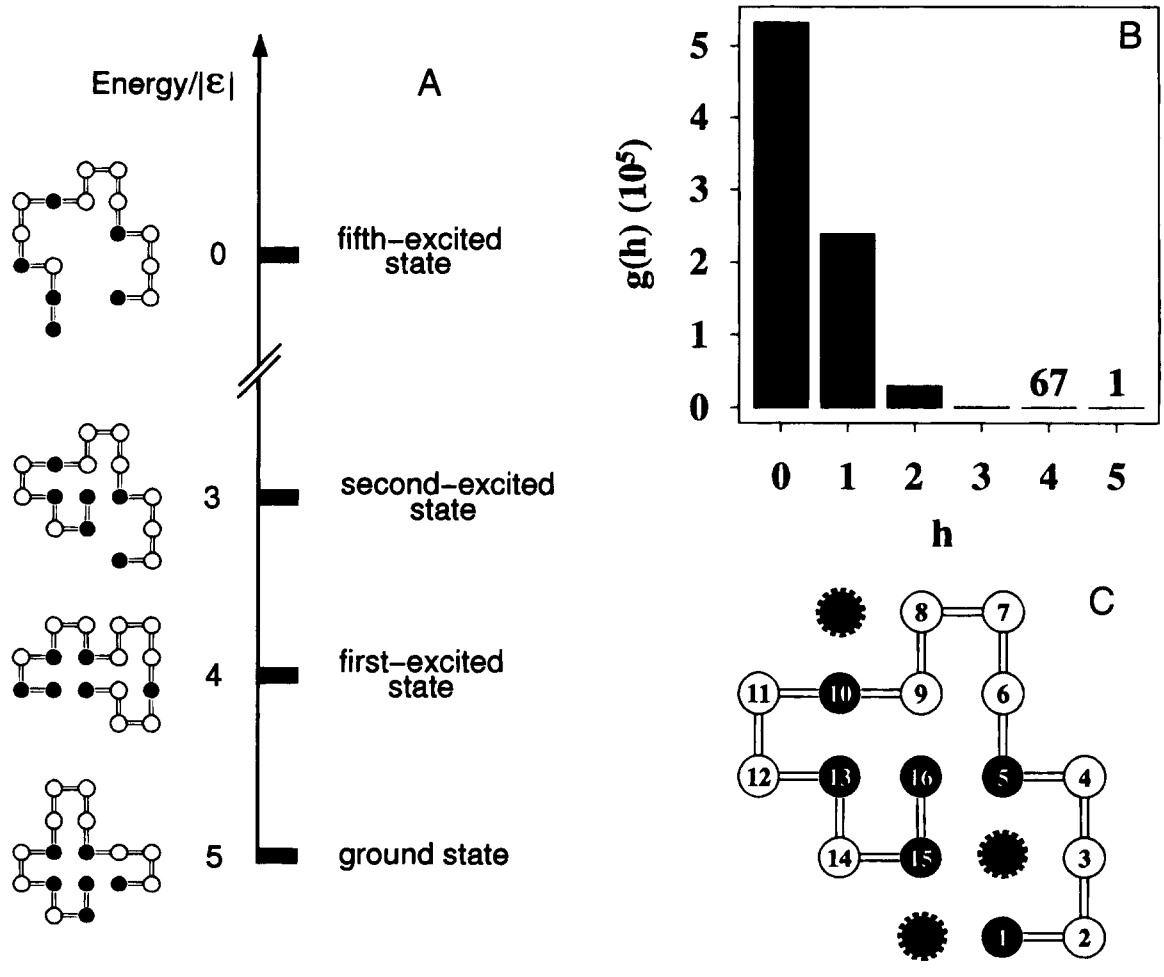


Fig. 1. **A:** Energy ladder diagram for 16-mer HP sequence A. Each hydrophobic (HH) contact is favored by a free energy ϵ. The native conformation, with five HH contacts, is at the bottom, and each step up the ladder represents the loss of one HH contact. **B:** Density of states, $g(h)$: the numbers of conformations of sequence A having h HH contacts. **C:** A particular ligation state with three ligands bound (protein monomers are numbered). Ligands are highlighted "beads." Contacts between a ligand and H monomer (LH contacts) are favored by a free energy $b\epsilon$, where b is a constant. This ligation state makes six LH contacts and three HH contacts, so the energy is $3\epsilon + 6b\epsilon$ (Equation 1).

provided that all possible ligation states can be enumerated. The HP model permits this, and thus provides a way to relate the binding thermodynamics to the structure and stability of the protein. Using Equation 1 for the ligation-state energy, E_s, of the HP model protein–ligand complex, Equation 2 becomes

$$P_s = \frac{e^{-h\epsilon/kT} e^{-mb\epsilon/kT} e^{\mu N_s/kT}}{\sum_{s=1}^{\Gamma} e^{-h\epsilon/kT} e^{-mb\epsilon/kT} e^{\mu N_s/kT}}. \quad (3)$$

Equation 3 can be computed exactly by exhaustive enumeration for any short HP sequence, thus providing average properties of model proteins as a function of temperature, ligand concentration, binding constants, and monomer sequence.

Despite its simplicity, the HP model offers several advantages for studying principles of protein–ligand interactions. (1) We can consider all model protein conformations, so there is no approximation or partial sampling of the protein conformational space, allowing a complete study of the effects of binding on protein structure. (2) We can compute all possible ligation states, for every possible chain conformation, so we are not limited to one or a few ligands bound at a time. We can explore a full range of ligand concentrations, from zero to denaturing. (3) The model has only two energy parameters, so we can explore the physics in a complete way. We make no assumptions about the locations or numbers of binding sites, about the mechanisms of ligand-induced conformational changes, or about how binding is affected by external conditions. Rather, these properties are derived from the theory.

Our aim is not to describe the biology of the various interactions, which may be quite complex, but rather to show how a diverse collection of binding phenomena can be understood through a simple unified picture that relies on only a few basic, physical concepts.

Validation of the model: Protein-like binding

The model shows a range of behaviors that mimic real protein–ligand interactions. These can be divided into two classes, which

we call specific and nonspecific binding. We refer to binding as specific when a ligand binds to high-affinity (two or more LH contact) sites on the protein, and when binding follows a site-specific (Michaelis–Menten) isotherm as described below. We call binding nonspecific when many ligands bind, either to the native or denatured states of the protein, and when the binding isotherms do not show site-specific thermodynamics.

Nonspecific binding behaviors of the model

Denaturation

High-affinity ligands (roughly $0.5 < b < 1.0$) at high concentration can induce unfolding in model proteins (see Fig. 2). As ligand concentration increases, the translational entropy of binding becomes more favorable, so LH contacts are favored at the expense of HH contacts, driving the protein to unfold. This model result is similar to protein denaturation by urea and guanidinium chloride. Other simple models of denaturation have been explored previously (Alonso & Dill, 1991; Thomas & Dill, 1993).

Dyes and weak solvents

Low-affinity ligands ($b < 0.2$) have little effect on model protein structure, even at high ligand concentrations. For these ligands, binding is sufficiently weak that breaking HH contacts to make LH contacts is always unfavorable. Thus, at high ligand concentrations, all hydrophobic sites on the protein surface become saturated, but the native structure remains intact. This model mimics the behavior of certain dyes and organic solvents (Allen et al., 1996; Mattos & Ringe, 1996) that bind, but do not perturb, the native structure.

ANS-like ligands

For an intermediate range of affinities in our model ($0.2 < b < 0.5$), ligands are too weak to induce full denaturation, but are strong enough to shift the native-denatured equilibrium. This class of ligands binds preferentially to the *compact denatured* states of the model proteins. As a result, the average number of bound ligands follows a bell-shaped curve (Fig. 3A) as external conditions are changed from native to denaturing. The maximum in the number of bound ligands occurs under the intermediate solvent conditions at which the compact denatured states are most stable. This behavior resembles that of ANS, a hydrophobic dye that binds preferentially to the molten globule states of proteins (Semisotnov et al., 1987, 1991; Shi et al., 1994) (see Fig. 3B).

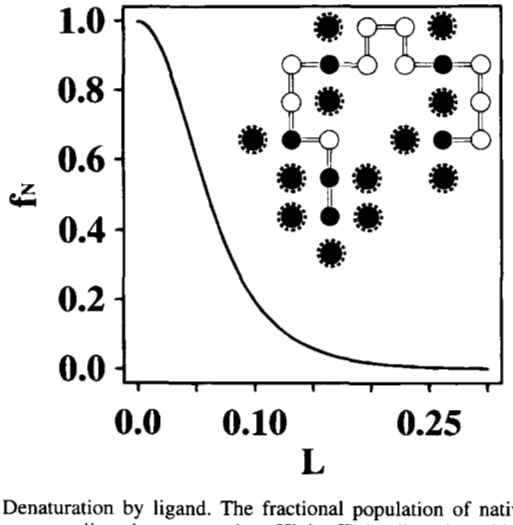


Fig. 2. Denaturation by ligand. The fractional population of native proteins, f_N, versus ligand concentration. High-affinity ligands at high concentrations denature model proteins. ($b = 0.50$, $\epsilon = -10$). Inset: Sample ligation state from the ensemble of denatured conformations.

Figure 3C explains this behavior. The number of ligands bound depends on a product of two factors: the number of accessible hydrophobic sites on the protein, and the strength of the binding interaction under the given conditions. Under native conditions (i.e., large $|\epsilon/KT|$), binding is strong (see Equation 1), but there are few available hydrophobic sites on the predominantly native protein, so there is little binding. Under denaturing conditions, the protein is unfolded and there are many exposed H sites, but binding is limited by the weak LH attraction. The number of bound ligands per protein is highest between these extremes, where the compact-denatured states are highly populated.

Semisotnov et al. (1987, 1991) have interpreted ANS binding as requiring "hydrophobic clustering" in the protein. In our model, hydrophobic clustering happens too, but clustering is a *consequence* of the balance between binding strength and number of binding sites, not a special mechanism of binding. In our model, there is nothing different about binding to a cluster than to any other arrangement of the same number of hydrophobic monomers.

Specific binding behaviors of the model

To define site-specific binding, we begin with the traditional mass-action description. Equation 4 illustrates the equilibrium between a native conformation, N, and its bound state, NL.

$$N + L \overset{K_{bN}}{\rightarrow} NL, \tag{4}$$

where K_{bN} is the binding equilibrium constant and L represents free ligand.

The fraction, f_b, of bound N molecules is (see Appendix)

$$f_b = \frac{K_{bN} L}{1 + K_{bN} L}, \tag{5}$$

which approaches unity as ligand concentration increases. We refer to behavior described by Equation 5 as site-specific, or "Michaelis–Menten" binding.

Lock-and-key and induced-fit binding

For some HP model proteins, and for high-affinity ligands ($0.5 < b < 1.0$), binding is localized to a single site on the protein and follows a Michaelis–Menten binding isotherm. Figure 4 shows a "lock and key" example of specific binding, in which the protein binds in its native conformation. The computed binding isotherm closely follows the Michaelis–Menten binding of Equation 5.

Site-specific binding might be considered surprising in this model, for two reasons. First, the model allows large numbers of alternate sites among the non-native and native states, because each H monomer is a potential contact. Second, the model protein–ligand interactions are *orientationally nonspecific*, lacking the geometric requirements of hydrogen bonding, for example. The specificity in our model arises instead from the ability of the protein to configure in a specific way, namely with a compatible pocket: the protein and ligand cannot mutually find any lower-energy configuration. Al-

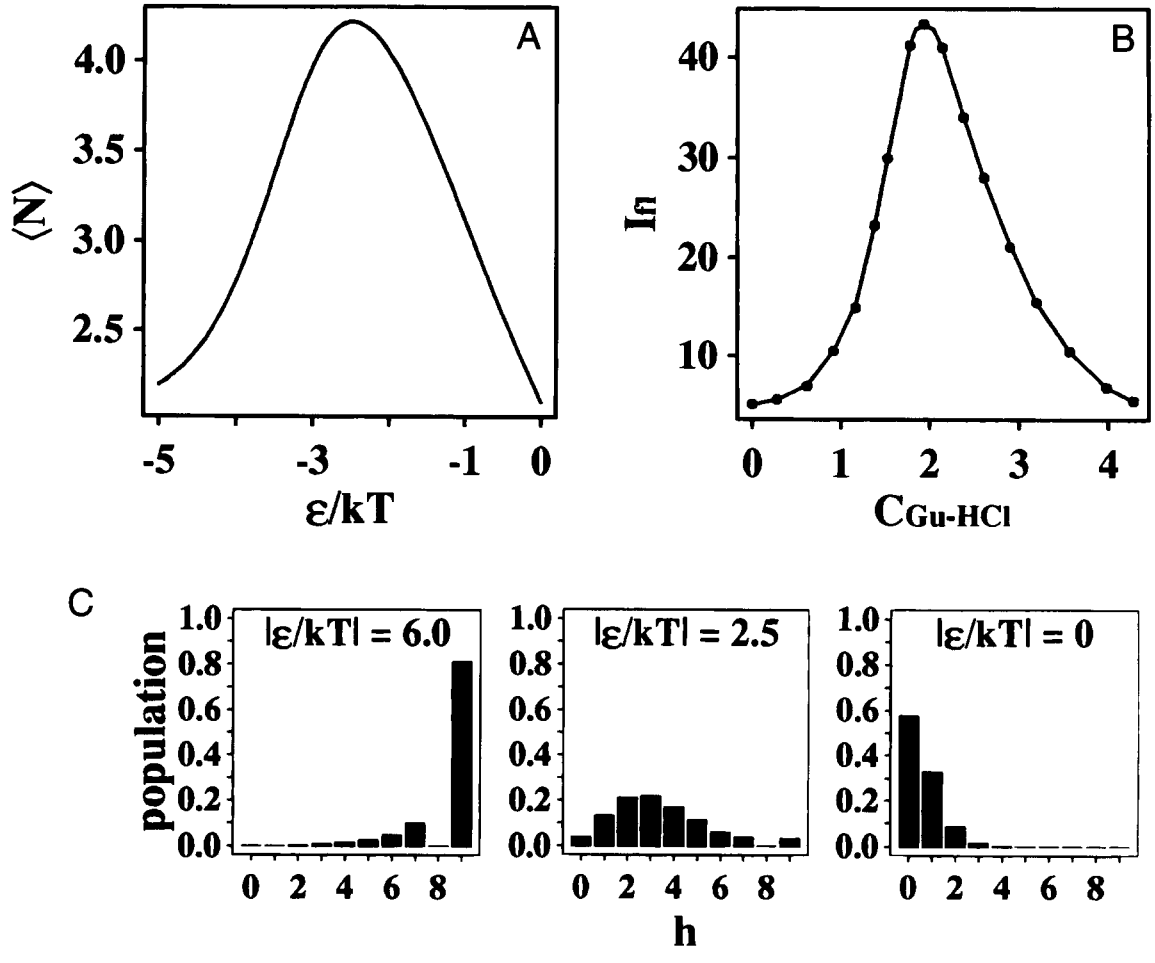


Fig. 3. Model of ANS binding. Ligands with *intermediate* affinities bind preferentially to compact-denatured states, much like ANS binds to molten globules. **A:** Average number of bound ligands versus external conditions ϵ/kT (native stabilizing conditions to the left; denaturing conditions to the right) with $L = 0.15$ and $b = 0.50$. **B:** Corresponding experiment: Fluorescence-intensity changes from ANS binding to bovine carbonic anhydrase B versus Gu-HCl concentration (adapted from Semisotnov et al., 1991). **C:** Energy-ladder histograms for the protein conformations, showing that, under conditions where binding is greatest ($\epsilon/kT = 2.5$ in A), the molecules have intermediate compactness ($h \approx 2$–4, middle of C).

though real proteins often take advantage of chemically specific interactions such as hydrogen bonds or salt bridges, our minimal model shows that binding specificity does not require it.

What is the basis for the specific binding in the model? There are two reasons the protein has relatively few binding options under native conditions. First, if a ligand is to bind a non-native conformation, the system must pay an energetic price to "excite" the protein from the native state to the non-native state. Even though the numbers of protein conformations and potential binding sites grow dramatically with increasing steps up the energy ladder, Boltzmann's law dictates that the ligand prefers to choose from the relatively few ligation states low on the energy ladder. Second, the binding of more than one ligand is disfavored by the high price in translational entropy at low ligand concentrations. We observe lock-and-key binding for approximately 11% of unique-folding 16-mer HP sequences. These are all sequences in which there are at least two H monomers exposed in the native state in the form of a model "binding pocket."

Our model also shows induced-fit binding, in which a ligand binds specifically a low-energy, but non-native, conformation of the protein. In these cases, the energy price in inducing the conformational change is more than compensated by the energy gain upon ligand binding. We observe induced fit in roughly 17% of HP sequences.

Modeling hemes and cofactors

For some HP sequences that do not fold to unique structures by themselves, a ligand can induce the "selection" of a single conformation (see Fig. 5). This is a model for proteins that populate a small conformational ensemble in the absence of cofactor, substrate, or prosthetic group, but that become structured in the bound complex. The heme-induced shift from apomyoglobin to myoglobin is an example.

Cooperative binding

Figure 6 shows an example of binding cooperativity between two identical ligands. Binding of the first ligand enhances the binding of the second, resulting in a cooperative binding isotherm. Depending on the HP sequence, binding can have different degrees of cooperativity and may or may not progress through a singly bound intermediate state.

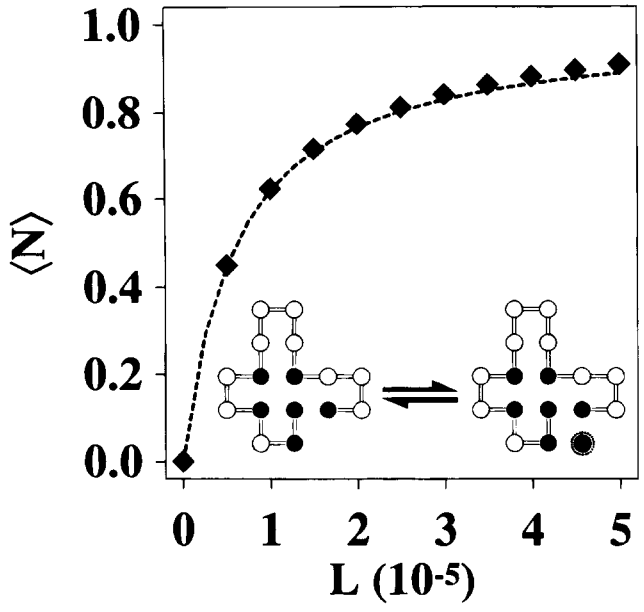


Fig. 4. Lock-and-key binding. The average number of bound ligands, $\langle N \rangle$ (diamonds), versus ligand concentration closely follows the theoretical curve (dashed line) expected for site-specific (Michaelis–Menten) binding ($\epsilon = -10$ and $b \approx 0.6$). Inset: The ligand binds at a single site on the native protein, without perturbing its structure.

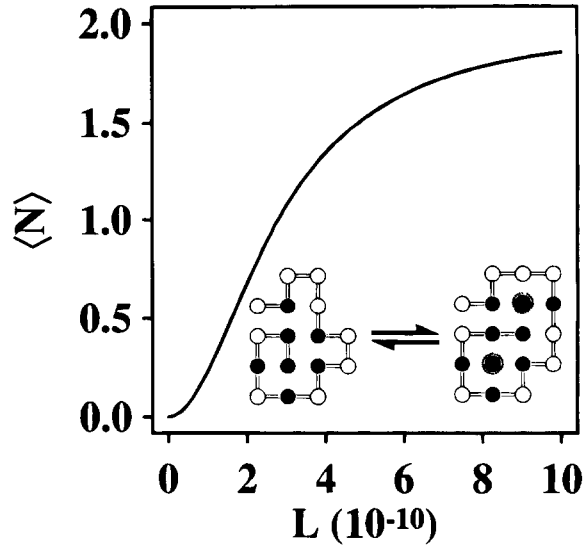


Fig. 6. Binding cooperativity. Average number of bound ligands, $\langle N \rangle$, versus ligand concentration. Binding of one ligand facilitates binding of the second. Inset: Unbound native state and unique excited state with two ligands bound.

Unconventional binding behavior in the model

Identically shaped ligands can bind in different modes

Figure 7 shows one of the most interesting results of the model. It bears on two standard premises: (1) the *steric premise*, that the binding mode is predominantly determined by the shape of the ligand, and (2) the *continuity premise*, that a small change in the structure of a ligand should lead to a small change in the structure of the bound complex. Figure 7 shows a case in which both of these premises are violated. Two identically shaped ligands bind in different locations depending on whether the binding is tight or weak. The tight binder ($b = 1$) binds to a second-excited state whereas the lower-affinity ligand ($b = 0.6$) binds to the native state in a lock-and-key fashion.

Because the two ligands have identical shapes, the choice of binding mode in this case is not based on shape complementarity alone, but also on the balance between the energy *lost* in inducing fit, and the energy *gained* in the binding. The tighter binder overcomes an unfavorable distortion of the protein because the resulting binding complex has a lower overall energy. This result resembles a recent experiment by Morton and Matthews (1995), which showed that ligands of very similar shape but different hydrophobicities can change the protein structure in different ways.

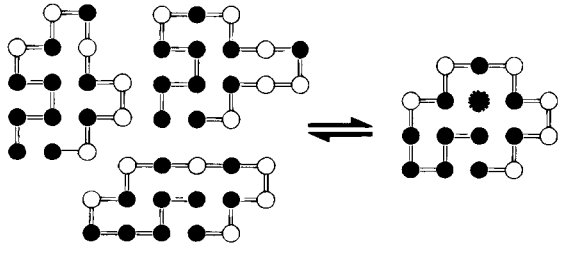


Fig. 5. Model heme binding. An HP sequence having 14 lowest-energy conformations (only three are shown) is locked into a single lowest-energy structure when the ligand binds.

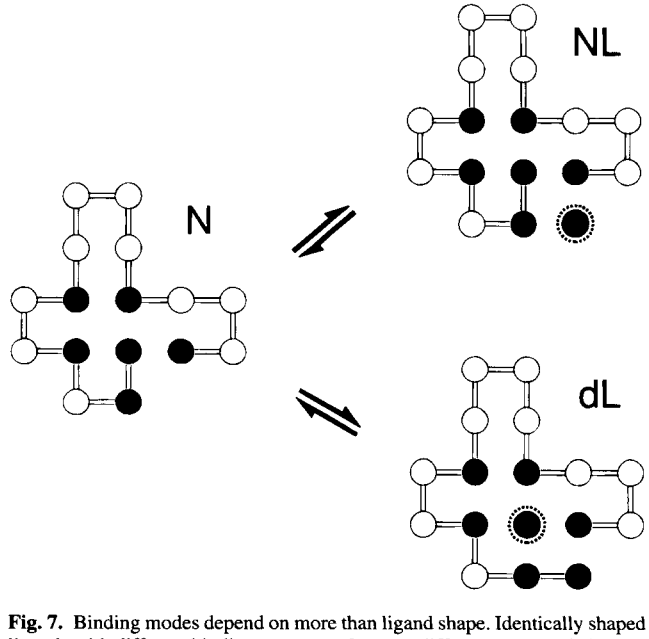


Fig. 7. Binding modes depend on more than ligand shape. Identically shaped ligands with different binding constants, b, cause different structural changes upon binding. When $b = 0.6$, binding favors the native-state complex, *NL* (lock and key), whereas $b = 1$ favors an induced-fit complex, *dL*, where *d* is a second-excited state. The low-energy complex is determined not by shape complementarity alone, because the two ligands have identical shapes, but by a balance between free energy *lost* in induced-fit structural changes, and free energy *gained* through the protein–ligand interaction.

We believe there are two implications for drug design. First, even small structural differences between ligands could lead to large differences in the binding mode or binding site. This could make the prediction of relative binding affinities sometimes difficult. Second, it implies that one possibly underappreciated determinant of the binding mode of a ligand is not structural. In Figure 7, the change in binding mode can be driven by temperature, or more generally by a change in solvent conditions, both of which regulate the binding strength. A ligand may choose its binding mode based on external conditions rather than on the shapes of the ligand or receptor, and might hop from one site to another when solvent conditions are changed.

Ligands can bind tightly to non-native states

Equation 4 describes Michaelis–Menten binding of a ligand to the native conformation, N. Now suppose a ligand may interact with either the native state or a particular non-native conformation, d. The corresponding mass-action equation becomes

$$\begin{array}{ccc} N + L & \overset{K_{bN}}{\rightarrow} & NL \\ \downarrow K'_u & & \downarrow \\ d + L & \overset{K_{bd}}{\rightarrow} & dL \end{array}, \tag{6}$$

where K_{bd} is the equilibrium binding constant for the non-native conformation, and K'_u is the equilibrium constant for N unfolding to d. We use the symbols d and D to represent non-native states: d is one particular non-native conformation (a microscopic state), and D represents the full ensemble of all non-native conformations (the macroscopic denatured state). The native state, too, in reality is an ensemble of microstates, but in the lattice model, we approximate N as a single microscopic conformation.

The fractional population of native conformations bound to ligand, f_b, is

$$f_b = \frac{K_{bN}L}{1 + K_{bN}L + K'_u K_{bd} L}. \tag{7}$$

When a ligand can bind d, the native state can never be fully saturated, because some ligand molecules bind to non-native conformations (NL and dL are assumed to be experimentally distinguishable). However, Equation 7 also shows that, even if a ligand binds the non-native state *more tightly* than it binds the native state (i.e., $K_{bd} > K_{bN}$), the effect on the binding curve, f_b, can be vanishingly small, provided that the non-native state has a very small population, $K'_u \ll 1$. Thus, tight binding doesn't necessarily imply high populations.

Two common objections to the notion of tight binding to non-native states are readily addressed: (1) How could a non-native conformation form a tighter binding site than the active site? The chain could "envelop" the ligand more completely than the native structure does, creating more LH contacts. (2) How could denatured-state binding overcome the unfavorable entropy required to restrict the presumably flexible unfolded conformation to a single, rigid structure in the bound complex? We distinguish between the *macroscopic* denatured state, D, and the large number of individual *microscopic* conformations, d, which comprise it. Any given denatured conformation d requires no more or less conformational entropy of binding than the native state does, because each is a single conformation. Hence, ligand binding to individual non-native conformations need not be intrinsically opposed by conformational entropy.

Although Equation 7 gives the mass-action scheme for non-native state binding, it gives no insight into the structural basis for this behavior. Figure 8A illustrates the structural basis using the HP model. The figure shows one example of a non-native state, d, of sequence A that binds ligand more tightly than the native state does, forming three LH contacts as opposed to only two for the native structure. Yet, as shown in Figure 4, the calculated binding curve indicates binding to only the native conformation. Why?

The reason is illustrated in Figure 8B, C, and D, which shows the corresponding free energy diagrams. Figure 8B shows the en-

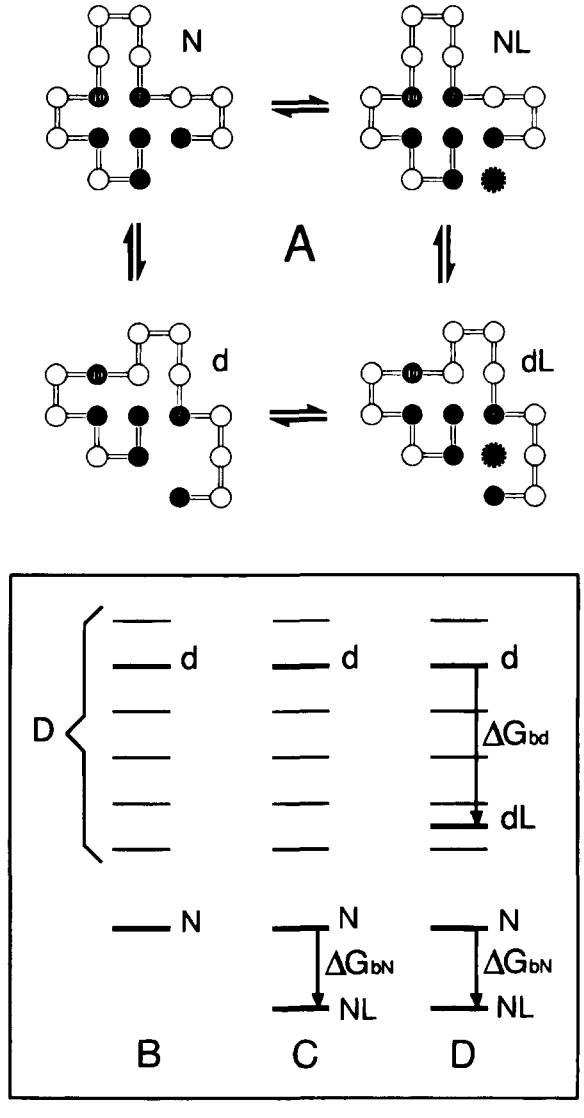


Fig. 8. Denatured-state binding. **A:** Ligand binds more tightly to a denatured (second-excited) state, d, than to N, because there are three LH contacts in the dL complex versus only two in the NL complex. But NL remains the dominant complex, and binding is Michaelis–Menten, because the cost of unfolding to d is large compared to the binding energy (see Fig. 4). **B:** Protein energy ladder, no ligand. **C:** Ligand binds only to the native state; the free energy of binding is ΔG_{bN}. **D:** Ligand binds d *more tightly* than N ($\Delta G_{bd} > \Delta G_{bN}$), as in A. Because the dL complex remains high in energy, it does not affect the binding isotherm.

ergy ladder of the protein conformations alone, in the absence of ligand. Figure 8C shows conventional binding, in which the ligand stabilizes the protein by binding the native state. Figure 8D shows the unconventional behavior represented in A: ligand stabilizes the native state, but it stabilizes a particular non-native conformation, *d*, even more. In this case, binding *increases* the population of *d*, but, because the *dL* conformation is so low in population relative to *NL*, the binding isotherm is insensitive to this non-native binding. Under the Michaelis–Menten conditions of Figure 4, all of the possible *dL* complexes of sequence A have a combined population of only 1×10^{-3} at saturating ligand concentration.

Hydrogen exchange can detect non-native binding

How can tight binding to low-population conformations be detected? Hydrogen exchange (HX) is a technique capable of detecting events of extremely low probabilities in proteins (Woodward & Hilton, 1979; Englander & Kallenbach, 1984; Miller & Dill, 1995; see Appendix for brief description). Although most experimental measurements of proteins give only ensemble averages, and are therefore overwhelmed by the native signal under native conditions, HX sees no contribution from the native state for protected hydrogens in the cores of proteins. In hydrogen exchange, the detection limit for low-population states is set by the longest exchange time that can be measured, which can be thousands of hours. Hence, HX routinely detects structures that are populated only sparsely and cannot be observed by other experimental methods.

When a ligand binds a protein, there are two ways it might affect the exchange rates of amide hydrogens from the hydrophobic core. (1) *Exchange from the unbound state.* If the ligand binds to only the native state (Fig. 8C), or at most to only a small number of non-native states, then then there will exist other non-native conformations that *cannot* be populated while the ligand is bound. For the amide hydrogens whose exchange is predominantly from these conformations, the ligand must dissociate transiently from the protein before exchange can occur. The free energy cost for this dissociation is equal to the binding free energy, so the resulting increase in HX rate is $\Delta\Delta G_{hx} = -\Delta G_b$ (see Appendix). For a typical protein, this represents an increase of several kcal/mol. (2) *Exchange from the bound state.* If the ligand *can* bind non-native conformations (Fig. 8D), there will be some amide hydrogens in the protein core that can exchange from *bound* protein structures. Because ligand binding will affect the population of these non-native "exchange" conformations, the corresponding HX rates will also change, by an amount equal to

$$\Delta\Delta G_{hx} = -kT \log \frac{K_{bd}}{K_{bN}} \tag{8}$$

(see Appendix), where K_{bd} and K_{bN} are the binding constants for the non-native and native states, respectively (note that K_{bN} approximates the experimentally measured binding constant, K_b, if *D* is low in population). Equation 8 can therefore be used to calculate the non-native state binding free energy for a given change in the HX rate.

Can HX detect strong interactions between a ligand and a highly non-native protein conformation? Our model predicts three experimental conditions that must be satisfied. (1) Exchange must be slow in the absence of ligand, i.e., $\Delta G_{hx} \approx \Delta G_u$ (see Appendix). This suggests that the exchange conformation is significantly non-native in structure. (2) Exchange must occur from the bound state, not the unbound state. This occurs when the change in HX rate is less than the binding energy, that is, when $\Delta\Delta G_{hx} < -\Delta G_b$. (3) The interaction of ligand with the exchange conformation must be strong. This is evidenced, through Equation 8, by a change in HX rate that is either negative or near zero, the former suggesting even *tighter* binding to the non-native than to the native state.

We use the HP model to simulate the effects of ligand binding on protein HX rates. In this model, the HX rate is proportional to the Boltzmann-averaged solvent accessibility of an amino acid over all possible protein conformations (Miller & Dill, 1995; see Appendix). In any one particular protein structure, a monomer is considered to exchange fully with solvent if it is adjacent to a solvent site. It is considered fully protected if it is surrounded on all four sides either by protein monomers or by ligand. Model HX rates are equal to the conformational ensemble average of this solvent accessibility quantity (either 0 or 1). For example, in Figure 8A, monomers 5 and 10 are buried in the native structure, and thus have much slower HX rates than the monomers exposed on the surface. But although monomer 5 is able to exchange from a *first*-excited state (see Fig. 1A), monomer 10 is buried in all of the first-excited states (not shown), and so can exchange only from a *second-* or higher-excited state. Hence, it exchanges much more slowly than monomer 5.

Figure 9A shows how ligand binding affects HX rates in the model. For monomer 5, the HX rate is decreased when the ligand binds, and the curve of HX rate versus ligand concentration indicates that exchange occurs predominantly from *unbound* conformations. This is because the first-excited states, the most important exchange conformations for monomer 5, do not bind the ligand as tightly as the native state does. First-excited states make only one LH contact, on average, versus two for the native state. Hence, for this monomer, the fastest route to exchange is by dissociation of the ligand, followed by unfolding of the protein. In contrast, the HX rate of monomer 10 *increases* upon ligand binding. As indicated in Figure 8A, this is due to tight binding to non-native states, particularly a small group of second-excited states. Using Equation 8, we find that the average binding constant, K_{bd}, for these second-excited state conformations is roughly 40 times that of the native state. Hence, binding constants can be predicted for extremely low-population *d* states using the measured change in HX rate (see further discussion in the Appendix).

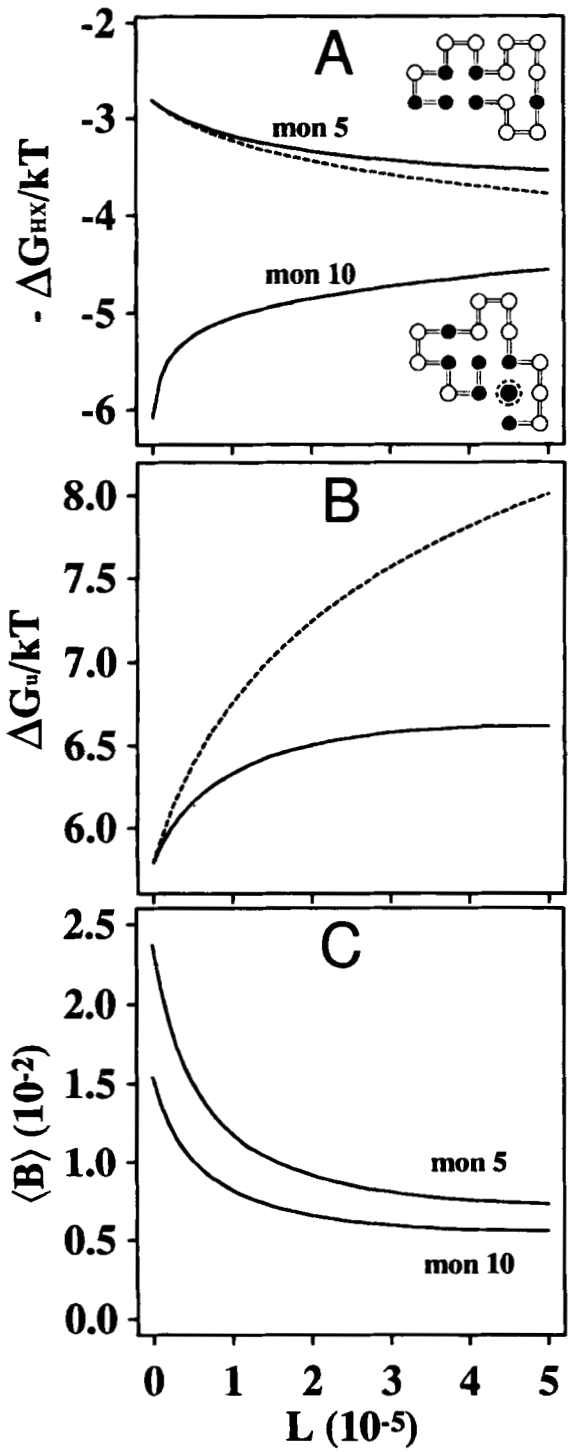
Experimental results from hydrogen exchange

There is experimental evidence from hydrogen exchange that ligands can bind tightly to highly non-native conformations. The HX technique has been used to study the effects of ligand binding on the structure and dynamics of cytochrome *c* (Paterson et al., 1990), lysozyme (Benjamin et al., 1992), serine-protease inhibitor (Werner & Wemmer, 1992), staph nuclease (Loh et al., 1993), barnase (Meiering et al., 1993), protein G (Orban et al., 1994), and acyl coenzyme A binding protein (ACBP) (Kragelund et al., 1995). In all of these proteins, the general effect of ligand binding is to decrease the HX rates of the majority of amide hydrogens. For many of these hydrogens, binding decreases the HX rate by the maximum amount, indicating that exchange occurs from the unbound states. Monomer 5 in Figure 9A shows this behavior.

However, the HX rates of many hydrogens either *increase* in the bound conformation, with a typical range in $\Delta\Delta G_{hx}$ of roughly -0.5 to -2.0 kcal/mol, or else they decrease by significantly *less* than the maximum amount. Either of these observations indicates

that exchange takes place from *bound* conformations. Among such hydrogens, many are among the *slowest* exchanging in the protein ($\Delta G_{hx} = \Delta G_u$), suggesting that exchange may take place from very non-native conformations. We believe these hydrogens are analogous to monomer 10 in Figures 8 and 9, and may be examples in which the ligand binds a largely unfolded protein conformation with *at least* as high an affinity as it binds to the native state.

Are there other explanations for this data? Several possible explanations have been ruled out. First, it is generally found that the increases in HX rate are not due to large structural changes upon binding. In all the examples cited above, there are only very small differences between the bound and unbound structures (from NMR or X-ray crystallography) near the relevant hydrogens. Second, there is no evidence that exchange is increased by bound water molecules trapped at the protein–ligand interface. Third, Benjamin et al. (1992) have argued on the basis of electrostatic calculations that ligand binding does not significantly affect the intrinsic chemical rate, k_x, of the exchange process (see Appendix). Finally, although it has been argued that HX might occur by "solvent penetration," rather than by structural unfolding (e.g., Woodward & Rosenberg, 1971; Lumry & Rosenberg, 1975; Richards, 1979), and thus that slow exchange might not imply highly unfolded exchange structures, HX experiments in urea (e.g., Hilton et al., 1981; Bai et al., 1994) suggest that, for the *slowest-exchanging* hydrogens in the protein core, exchange involves considerable unfolding of the protein.



Other explanations for increased HX rates upon ligand binding can be found in the hydrogen exchange literature. These involve either: (1) changes in protein structure too subtle to be detected with the measurement techniques used (NMR and X-ray crystallography); (2) a ligand-induced decrease in the "regional stability" of the protein (Kragelund et al., 1995); or (3) an increase in protein dynamics (Benjamin et al., 1992; Kragelund et al., 1995). But there are problems with each of these explanations. For (1), rate increases may be due to subtle changes in *structure* (Equation 7), but they cannot be due to subtle changes in the binding *energetics*. Our model predicts that in order for hydrogens to undergo even a *small* increase in HX rate, the ligand must bind the non-native exchange conformation *at least* as tightly as it binds the native state. The problem with explanation (2) is only that "regional stability" is not very clearly defined. Moreover, *global* stability has been observed to *increase* even when the HX rates of some hydrogens also increase (Loh et al., 1993; Kragelund et al., 1995). Similar behavior arises from our model, as shown below. Explanation (3) appears inconsistent with the poor correlation between changes in HX rates and changes in protein flexibility (Kragelund et al., 1995, using T_1 times). As shown below, our model too, predicts a poor correlation.

Our model results are fundamentally different from the previous proposals listed above. Our explanation puts less emphasis on the native state alone, and more emphasis on ligand binding to non-native conformations. Non-native state binding can account for affects on HX rates even when there is little or no detectable change in the protein's native structure.

Fig. 9. Effects of binding on HX rates, stability, and flexibility. **A:** Binding *increases* the HX rate of monomer 10 by stabilizing *second*-excited states (one shown), which are the fastest-exchanging states for this monomer. Binding *decreases* the HX rate of monomer 5 by stabilizing the *native* state relative to the weak-binding *first*-excited states (one shown), which are the fastest-exchanging states for monomer 5. The HX rate of monomer 5 is close to the theoretical rate (dashed line) corresponding to HX from only the *unbound* states. **B:** Binding stabilizes the native state (increases ΔG_u) because binding is weaker to the full denatured ensemble than to the native state. Dashed line is the expected stability change when *only* the native state is bound. The actual stabilization is smaller because of binding to denatured states. **C:** Average thermal factors, $\langle B \rangle$, decrease with ligand binding for both monomers 5 and 10. The thermal motions depend mainly on first-excited states, which bind ligand weakly, whereas the HX of monomer 10 depends on second-excited states. For A–C, the model parameters are identical to Figure 4.

How can ligand binding stabilize the native state while increasing the HX rate?

Figure 9B shows that a bound ligand stabilizes the native state of sequence A, while at the same time increasing the HX rate of monomer 10 (Fig. 9A). This surprising result has a simple explanation: the ligand affects some few non-native conformations, d_1, d_2, d_3,..., differently than it affects the overall denatured ensemble, D. In particular, the non-native conformations contributing most to the D ensemble, and thus to protein stability, are the first-excited states, which for this HP sequence bind the ligand less favorably than the native state does. On the other hand, the most important conformations for the HX of monomer 10 are second-excited states, many of which bind ligand *more tightly* than the native state does (Fig. 8A). These results are in agreement with experiments from ACBP (Kragelund et al., 1995), which show that both global stability and HX rates can be increased simultaneously by ligand binding, and from early experiments by Woodward (Hilton et al., 1981) showing that HX rates and global stability can be poorly correlated.

Michaelis–Menten binding can sometimes be destabilizing

Figure 10 illustrates another unexpected result: that favorable ligand binding can be globally *destabilizing*. Figure 10A shows the binding curve of a particular HP sequence that appears to bind a ligand in a lock-and-key fashion to only the native conformation. Figure 10B shows that the same binding event *decreases* the global stability. The reason is that ligand interacts strongly with conformations of the denatured ensemble, thereby increasing the denatured-state population relative to the native. The difference between this and the previous example, where binding *increased* stability, is that here the individual dL conformations *are* highly representative of the full D ensemble, because they are among the first-excited states. Nevertheless, their population *relative to the native state* remains very small. Thus, although the fraction of molecules that are denatured changes several fold upon binding, causing a significant decrease in global stability, the fraction itself remains too small to affect the binding isotherm (i.e., K'_u in Equation 7). This result suggests that observing Michaelis–Menten binding does not necessarily rule out the possibility of highly non-native protein configurations in the bound complex. A possible example of this behavior is the protein ACBP (Kragelund et al., 1995), which shows an increase in stability much smaller than expected given the strength of the interaction with its inhibitor (see Appendix).

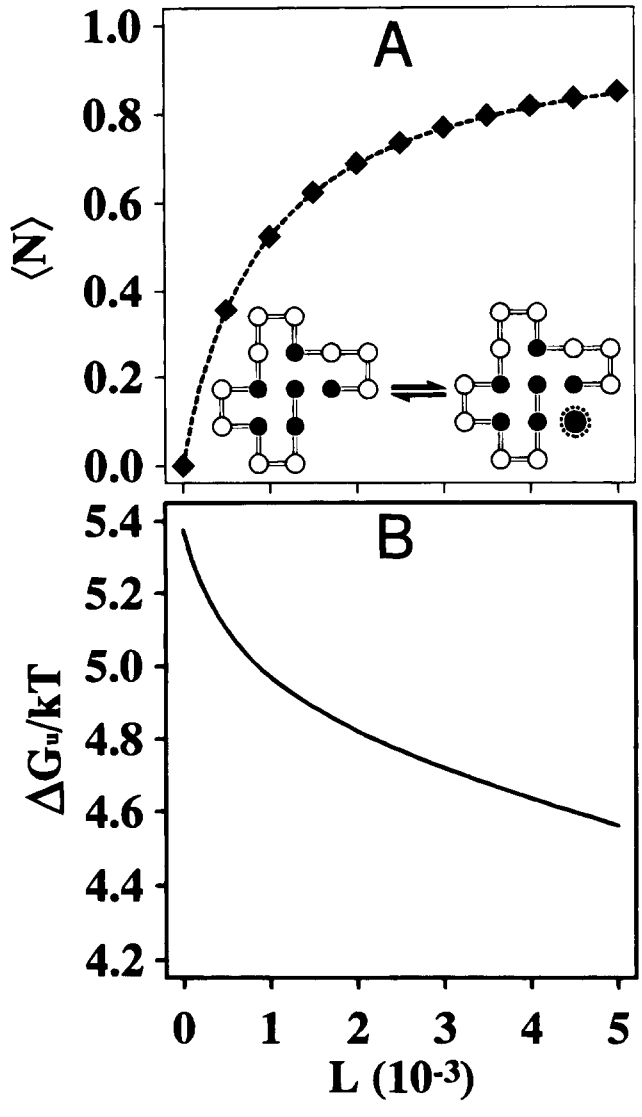


Fig. 10. Michaelis–Menten binding can be globally destabilizing. **A:** Sample HP sequence shows lock-and-key, Michaelis–Menten binding (same as Fig. 4). **B:** Ligand binding *decreases* the free energy of unfolding, ΔG_u, by stabilizing first-excited states (not shown), and thus increasing the D population relative to N. However, because the overall D population remains small (less than 1% in this example), binding to d states has only a small effect on the binding curve ($b = 0.35$ and $\epsilon = -10$).

HX rates do not always correlate with protein flexibility

Figure 9C shows that ligand binding can restrict the thermal motions of two residues, even though one residue has an increased HX rate and the other has a decreased HX rate. We calculate the thermal motion in the model as the Boltzmann-average of the fluctuations over all the intra-monomer distances, resembling a B factor in X-ray crystallography (see Appendix). B factors are often taken as a measure of the flexibility of the protein, and our model, like experiments, shows a general correlation between thermal motions and HX rates. However, this is not true of the *changes* that occur in these measurements as result of ligand binding. Figure 9C shows that, although binding *increases* the HX rate of monomer 10, it *decreases* the B factor. This result has the same explanation as above: B factors reflect ensemble averages over the first-excited states, whereas the HX of monomer 10 reflects binding to the small subpopulation of second-excited states. Although there is often at least some correlation observed between absolute HX rates and protein dynamics, such as in BPTI (Levitt, 1981), ribonuclease A (Wlodawer & Sjolin, 1982), and trypsin (Kossiakoff, 1982), there is at least one observation in which ligand binding *increases* the HX rate while simultaneously *decreasing* flexibility (Kragelund et al., 1995, with ACBP using T_1 times), consistent with the model result shown in Figure 9C.

We conclude that HX rates and thermal B factors may not always reflect the same property of a protein. For the slowest-exchanging hydrogens in particular, the structures that determine HX rates are higher up the energy ladder, whereas the structures that determine the flexibility are on the first rung. If the ligand can interact differently with the two classes of conformations, then our

model suggests that changes in HX rates will not correlate well with changes in flexibility.

Discussion

Proteins are flexible and only marginally stable. But current models of ligand–protein interactions often assume that proteins are rigid and unperturbable, and that there is a single dominant mode of binding to the native structure. To test the validity of these ideas requires models that go beyond binding polynomials, so that structural consequences of binding are *derived* from principles and structures, rather than *assumed*. Here we present a first step in this direction. Using the two-dimensional HP lattice model, we search exhaustively through all of protein conformational space, and exhaustively through all possible ligation states of each conformation, to find the global free-energy minimum. In this way, we *compute* the structures and thermodynamics of binding from the protein sequence and the ligand structure, rather than postulating them.

Despite the simplicity of this model, it shows a wide range of protein–ligand binding behaviors. Very hydrophobic ligands at high concentrations denature model proteins, much like urea and guanidine hydrochloride. Weaker ligands bind to compact denatured states, modeling the behavior of ANS. Even weaker ligands bind native states, as dyes and probes do, without perturbing the protein structure. Model ligands can also bind specifically, that is, at single specific sites, and with Michaelis–Menten (Langmuir) isotherms. Such binding specificity does not require chemical specificity of the underlying interactions, such as hydrogen bonding or ion pairing. The model shows that such binding can be lock-and-key, to a preformed site on the native structure, or induced fit, to an excited-state conformation having a higher energy than the native state. We also find examples of a ligand stabilizing a disordered ensemble, much like heme stabilizes apomyoglobin, as well as examples of binding cooperativity between two ligands.

The model is useful for explaining nontraditional binding behaviors, many of which have been observed experimentally. Two ligands of identical shape can bind in very different binding modes. Ligands can bind denatured states more tightly than the native state and still show Michaelis–Menten binding isotherms. Binding that appears to be "lock and key" may actually destabilize the native structure. Binding may cause model HX rates to increase, while at the same time increasing global stability and decreasing thermal *B* factors. The model suggests that many of these nontraditional behaviors may have the same physical origin, namely that a ligand can interact differently with a few non-native states than it does with the vast sea of other denatured conformations in general. Experimental evidence of denatured-state binding is found in HX data from barnase, lysozyme, and ACBP.

There are two broader implications for understanding protein–ligand interactions. First, this model suggests that the premise of structure-based drug design—that knowledge of the native structure of a protein is sufficient to rationalize a binding interaction—may not always be true. Knowledge of non-native structures may sometimes be necessary. It suggests that certain information about ligand binding may be hidden and unavailable for structural interpretation by X-ray crystallography and NMR spectroscopy. Second, in the same way that the modeling of protein folding has moved from simple macroscopic mass-action models to more microscopic models that are rooted in the language of ensembles, energy ladders, and energy landscapes (Dill & Chan, 1997), ligand binding too can benefit from more microscopic models and from the language of energy landscapes. Most importantly, proteins are not single conformations. Even under native conditions, proteins populate a broad ensemble of structures. Some of these fluctuations may be important for binding, and may not be well represented by the average properties of the denatured ensemble.

Acknowledgments

We thank NIH for financial support. D.W.M. has received support from National Science Foundation Fellowship GER-9255688 (96-97).

References

Ackers GK, Doyle ML, Myers D, Daugherty MA. 1992. Molecular code for cooperativity in hemoglobin. *Science 255*:54–63.

Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D. 1996. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem 100*:2605–2611.

Alonso DOV, Dill KA. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry 30*:5974–5985.

Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins Struct Funct Genet 20*:4–14.

Benjamin DC, Williams DC, Smith-Gill SJ, Rule GS. 1992. Long-range changes in a protein antigen due to antigen–antibody interaction. *Biochemistry 31*:9539–9545.

Bohacek RS, McMartin C. 1994. Multiple highly diverse structures complementary to enzyme binding sites: Results of extensive application of a de novo design method incorporating combinatorial growth. *J Am Chem Soc 116*:5560–5571.

Chan HS, Dill KA. 1989. Compact polymers. *Macromolecules 22*:4559–4573.

Chan HS, Dill KA. 1990. Origins of structure in globular proteins. *Proc Natl Acad Sci USA 87*:6388–6392.

Chan HS, Dill KA. 1991. "Sequence space soup" of proteins and copolymers. *J Chem Phys 95*:3775–3787.

Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys 100*:9238–9257.

Creighton TJ. 1993. *Proteins: Structures and molecular properties, 2nd ed.* New York: W.H. Freeman and Company. pp 339.

Di Cera E. 1995. *Thermodynamic theory of site-specific binding processes in biological macromolecules*. Cambridge, England: Cambridge University Press.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nature Struct Biol 4*:10–19.

Englander SW, Kallenbach NR. 1984. Structural dynamics of proteins and nucleic acids. *Q Rev Biophys 16*:521–655.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science 254*:1598–1603.

Hilton BD, Trudeau K, Woodward CK. 1981. Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. *Biochemistry 20*:4697–4703.

Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv Protein Chem 21*:287–386.

Koshland DE, Nemethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry 5*:365–385.

Kossiakoff AA. 1982. Protein dynamics investigated by the neutron diffraction–hydrogen exchange technique. *Nature 296*:713–721.

Kragelund BB, Knudsen J, Poulsen FM. 1995. Local perturbations by ligand binding of hydrogen deuterium exchange kinetics in a four-helix bundle protein, acyl coenzyme A binding protein (ACBP). *J Mol Biol 250*:695–706.

Kuntz ID. 1992. Structure-based strategies for drug design and discovery. *Science 257*:1078–1082.

Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules 22*:3986–3997.

Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA 87*:638–642.

Levitt M. 1981. Molecular dynamics of hydrogen bonds in bovine pancreatic trypsin inhibitor protein. *Nature 294*:379–380.

Loh SN, Prehoda KE, Wang J, Markley JL. 1993. Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry 32*:11022–11028.

Lumry R, Rosenberg A. 1975. The mobile defect hypothesis of protein function. *Colloq Int CNRS 246*:55–63.
Mattos C, Ringe D. 1996. Locating and characterizing binding sites on proteins. *Nature Biotech 14*:595–599.
Meiering EM, Bycroft M, Lubienski MJ, Fersht AR. 1993. Structure and dynamics of barnase complexed with 3′-GMP studied by NMR spectroscopy. *Biochemistry 32*:10975–10987.
Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Sci 4*:1860–1873.
Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol 12*:88–118.
Morton A, Matthews BW. 1995. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: Linkage of dynamics and structural plasticity. *Biochemistry 34*:8576–8588.
Orban J, Alexander P, Bryan P. 1994. Hydrogen-deuterium exchange in the free and immunoglobulin G-bound protein G B domain. *Biochemistry 33*:5702–5710.
Paterson Y, Englander SW, Roder H. 1990. An antibody binding site on cytochrome *c* defined by hydrogen exchange and two-dimensional NMR. *Science 249*:755–759.
Richards FM. 1979. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. *Carlsberg Res Commun 44*:47–63.
Roder H, Wagner G, Wuthrich K. 1985. Amide proton exchange in proteins by EX1 kinetics: Studies of the basic pancreatic trypsin inhibitor at variable p2H and temperature. *Biochemistry 24*:7396–7407.
Schellman JA. 1975. Macromolecular binding. *Biopolymers 14*:999–1018.
Segal DM, Harrington WF. 1967. The tritium-hydrogen exchange of myosin and its proteolytic fragments. *Biochemistry 6*:768–787.
Semisotnov GV, Rodionova NA, Kutyshenko VP, Elbert B, Blank J, Ptitsyn OB. 1987. Sequential mechanism of refolding of carbonic anhydrase B. *FEBS Lett 224*:9–13.
Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*:119–128.
Shi L, Palleros DR, Fink AL. 1994. Protein conformational change induced by 1, 1′-Bis (4-anilino-5-naphthalenesulfonic acid): Preferential binding to the molten globule of DnaK. *Biochemistry 33*:7536–7546.
Shoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. 1993. Structure-based discovery of inhibitors of thymidylate synthase. *Science 259*:1445–1450.
Shortle D, Chan HS, Dill KA. 1991. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci 1*:201–215.
Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MNG. 1996. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nature Struct Biol 3*:233–239.
Thomas PD, Dill KA. 1993. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci 2*:2050–2065.
Werner MH, Wemmer DE. 1992. Identification of a protein-binding surface by differential amide hydrogen-exchange measurements. Application to Bowman-Birk serine-protease inhibitor. *J Mol Biol 225*:873–889.
Wlodawer A, Sjolin L. 1982. Hydrogen exchange in ribonuclease A: Neutron diffraction study. *Proc Natl Acad Sci USA 79*:1418–1422.
Woodward CK, Hilton BD. 1979. Hydrogen exchange kinetics and internal motions in proteins and nucleic acids. *Annu Rev Biophys Bioeng 8*:99–127.
Woodward CK, Rosenberg A. 1971. Studies of hydrogen exchange in proteins. VI. Urea effects on RNase hydrogen exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *J Biol Chem 246*:4114–4121.
Wyman J, Gill SJ. 1990. *Binding and linkage. Functional chemistry of biological macromolecules*. Mill Valley, California: University Science Books.

Appendix

Derivation of the partition function

Here we calculate the total partition function for our model of ligand binding, given by the denominator of Equation 3. This requires summing the individual Boltzmann factors for all possible combinations of ligand and model protein. We use "ligation state" to denote a particular combination of ligand and model protein, and "conformation" to denote the HP chain structure alone. Each chain *conformation* has many *ligation states*. A "binding site" is any vacant site adjacent to an H monomer. For a given HP sequence in a given conformation, the total number of ligation states is found by counting the number of ways that indistinguishable ligands can be distributed among the binding sites, beginning with zero ligands and ending with the number that fills all of the sites of the particular conformation.

First, we calculate the number of ways that ligands can bind to one particular conformation, *c*. We define four numbers, $M_i = 1, 2, 3, 4$, equal to the numbers of binding sites that contact i monomers of type H. The chain conformation in Figure 1C, for example, has $M_1 = 2$, $M_2 = 1$, $M_3 = 1$, and $M_4 = 0$. Now, for a particular *ligation state*, s, the number of ligands bound to each of the four types of sites is denoted by n_i, and can vary from 0 to M_i. The particular ligation state in Figure 1C has $n_1 = 1$, $n_2 = 1$, $n_3 = 1$, and $n_4 = 0$. The total number of bound ligands, N_s, in the ligation state is obtained by adding the four n_i values:

$$N_s = \sum_{i=1}^{4} n_i. \tag{9}$$

There is more than one way that n_i ligands can be distributed among the M_i sites. The total number of combinations, g_i, is given by the expression

$$g_i = \binom{M_i}{n_i} = \frac{M_i!}{n_i!(M_i - n_i)!}. \tag{10}$$

Similarly, the number of ways that all N_s ligands can be arranged such that n_1 of them bind among the M_1 single-contact sites, n_2 of them bind among the M_2 two-contact sites, and so on, is given by the product of four terms:

$$g_{total} = \binom{M_1}{n_1}\binom{M_2}{n_2}\binom{M_3}{n_3}\binom{M_4}{n_4} = \prod_{i=1}^{4}\binom{M_i}{n_i}. \tag{11}$$

The partial partition function, Ξ_c, for the given HP conformation, *c*, is obtained by summing the Boltzmann factors for each possible ligation state over all values of n_i

$$\Xi_c = \sum_{n_1=0}^{M_1}\sum_{n_2=0}^{M_2}\sum_{n_3=0}^{M_3}\sum_{n_4=0}^{M_4}\left(\left(\prod_{i=1}^{4}\binom{M_i}{n_i}\right)e^{-h\epsilon/kT}e^{-mb\epsilon/kT}e^{\mu N_s/kT}\right), \tag{12}$$

where h is the number of HH contacts for the conformation; N_s is the total number of ligands bound in ligation state s (Equation 9); and m is the total number of ligand contacts in the ligation state:

$$m = n_1 + 2n_2 + 3n_3 + 4n_4 = \sum_{i=1}^{4} in_i. \tag{13}$$

The total partition function, Ξ, is found by adding up the partial partition functions for all Ω possible HP chain conformations:

$$\Xi = \sum_{c=1}^{\Omega} \Xi_c, \tag{14}$$

where $\Omega = 802{,}075$ in the case of a 16-mer.

Comparison with the binding polynomial

Our microscopic binding partition function (Equation 14) can be related directly to the thermodynamic binding polynomial, Q. The binding polynomial is related to the free energy of binding, ΔG_b, by (Schellman, 1975)

$$\Delta G_b = -kT \log Q. \tag{15}$$

To compare the partition function with the binding polynomial, it is useful to express Equation 14 in a different form. Written as a product of four sums, Equation 14 becomes

$$\Xi = \sum_{c=1}^{\Omega} e^{-h\epsilon/kT} \prod_{i=1}^{4} \left\{ \sum_{n_i=0}^{M_i} \binom{M_i}{n_i} (e^{-ib\epsilon/kT} e^{\mu/kT})^{n_i} \right\}. \quad (16)$$

The analogous binding polynomial, Q, for macromolecular binding has been derived by Schellman (1975) and is given by

$$Q = \sum_{c=1}^{\Omega} K_c \prod_{i=1}^{4} \left\{ \sum_{n_i=0}^{M_i} \binom{M_i}{n_i} (K_{bi} L)^{n_i} \right\}, \quad (17)$$

where Ω is the number of molecular conformations, c; K_c are the equilibrium constants with respect to the $c = 1$ conformation ($K_1 = 1$); K_{bi} are the binding constants describing the four types, i, of binding site; M_i are the number of each type of site; n_i are the number of ligands bound to each type, and L is free ligand concentration.

Comparison of Equations 16 and 17 show the relationships between the microscopic Boltzmann factors and the corresponding thermodynamic quantities:

$$K_c = e^{-\Delta h\epsilon/kT} \quad (18)$$

$$K_{bi} = e^{-ib\epsilon/kT} \quad (19)$$

$$L = e^{\mu/kT}. \quad (20)$$

Hydrogen exchange

We assume that HX of amide hydrogens from the protein interior is governed by the equilibrium unfolding of the native structure (Englander & Kallenbach, 1984):

$$N(H) \underset{k_f}{\overset{k_u}{\rightleftharpoons}} d(H) \xrightarrow{k_x} d(D), \quad (21)$$

where $N(H)$ is the native structure; $d(H)$ and $d(D)$ describe an individual denatured conformation prior to and after the exchange of a hydrogen for a deuterium; k_u and k_f are the rate constants for partial unfolding and folding of the native structure; and k_x is the exchange rate for a nonbonded, solvent-exposed hydrogen. The observed rate constant, k_{obs}, for this process is (Hvidt & Nielsen, 1966; Segal & Harrington, 1967; Roder et al., 1985)

$$k_{obs} = Kk_x, \quad (22)$$

where $K = k_u/k_f$ is the equilibrium constant for the unfolding process. The observed rate is the product of the "maximum" rate, k_x, with K, the fraction of time the protein is in the exchanging conformation, d. Equation 22 can also be expressed as ΔG_{hx}, the free energy required for the protein to undergo the transition from the native to the exchanging conformation:

$$\Delta G_{hx} = -kT \log K = -kT \log \frac{k_{obs}}{k_x}. \quad (23)$$

This free energy is close to zero for surface hydrogens exchanging near the theoretical limit, k_x, and reaches a maximum roughly equal to the free energy of global unfolding, ΔG_u, for those hydrogens in the hydrophobic core that require complete unfolding for exchange.

In the HP lattice model, HX rates are calculated from Boltzmann-averaged solvent accessibilities, A. $A = 0$ if a monomer is fully protected, i.e., completely surrounded by other monomers or ligands, and $A = 1$ if the monomer is solvent accessible, i.e., is adjacent to an empty lattice site. The HX rate from the monomer is

$$k = \langle A \rangle k_x, \quad (24)$$

which is identical to Equation 22, except that the two-state equilibrium constant is replaced by the ensemble-averaged accessibility, $\langle A \rangle$.

Binding to non-native states

The effects of non-native state binding on global stability, hydrogen exchange, and the binding isotherm can be considered using the simple thermodynamic binding equilibrium:

$$\begin{array}{ccc} N + L & \overset{K_{bN}}{\rightarrow} & NL \\ \downarrow K'_u & & \downarrow \\ d + L & \overset{K_{bd}}{\rightarrow} & dL \end{array}, \quad (25)$$

where N and d represent the native state and a particular denatured conformation, respectively; K_{bN} and K_{bd} are the binding constants for N and d; and K'_u is the equilibrium constant for N unfolding to d.

If the ligand can bind to the native state, N, but not to d (i.e., if $K_{bd} = 0$), it follows that

1. The fractional population of bound native-state molecules, f_b, as a function of ligand concentration is

$$f_b = \frac{NL}{N + d + NL} = \frac{K_{bN} L}{1 + K_{bN} L}, \quad (26)$$

where we have assumed $K'_u \ll 1$ in the last step. This is the equation for Michaelis–Menten binding, and the denominator is equal to the Michaelis–Menten binding polynomial, Q.

2. The corresponding change in global stability upon ligand binding, i.e., the change in the free energy, $\Delta\Delta G_u$, is

$$\Delta\Delta G_u = -kT \log \frac{d}{N + NL} + kT \log \frac{d}{N} = -kT \log \frac{1}{1 + K_{bN} L}, \quad (27)$$

which is simply the free energy of binding, $-\Delta G_b$, from Equation 15, because $Q = 1 + K_{bN}L$ is the binding polynomial for Michaelis–Menten binding. Thus, if ligand binds to only the native state, all of the binding energy contributes toward stabilizing the native relative to the unfolded states.

3. The change in HX rate, $\Delta\Delta G_{hx}$, of a hydrogen whose fastest route to HX is through conformation d is given by an expression identical to Equation 27:

$$\Delta\Delta G_{hx} = -kT \log \frac{1}{1 + K_{bN} L} = -\Delta G_b. \quad (28)$$

Hence, if ligand binds only the native state, the HX rate is expected to decrease by an amount equal to the binding energy, $-\Delta G_b$, which is typically several kcal/mol. This is true for *all* hydrogens, regardless of the particular fluctuation d required for exchange.

Conversely, if the ligand can bind to the unfolded state, d, with a binding constant K_{bd}, it follows that

1. The fractional population of bound native molecules, f_b, becomes

$$f_b = \frac{NL}{N + d + NL + dL} = \frac{K_{bN} L}{1 + K'_u + K_{bN} L + K'_u K_{bd} L}. \quad (29)$$

Although f_b in Equation 29 is less than for the Michaelis–Menten case, it reduces to normal Michaelis–Menten binding when the unfolded-state population is very small ($K'_u \ll 1$), even when binding is *stronger* to d than to N ($K_{bd} > K_{bN}$).

2. The change in global stability will deviate from the Michaelis–Menten case above only if the *single* bound conformation d represents a significant population of the *full* denatured ensemble, D. In this case, the change in stability is

$$\Delta\Delta G_u = -kT \log \frac{d + dL}{N + NL} + kT \log \frac{d}{N}$$
$$= -kT \log \frac{1 + K_{bd}L}{1 + K_{bN}L}. \quad (30)$$

Equation 30 implies the well-known result that ligands that prefer to bind the native state will stabilize the protein, whereas ligands that prefer to bind D will destabilize the protein.

3. If hydrogens can exchange from the bound state, dL, then the change in HX rate is given by

$$\Delta\Delta G_{hx} = -kT \log \frac{d + dL}{N + NL} + kT \log \frac{d}{N}$$
$$= -kT \log \frac{1 + K_{bd}L}{1 + K_{bN}L}. \quad (31)$$

This is equal to the stability change of Equation 30, but Equation 31 is valid for *any* denatured conformation, d, no matter how low in population. At saturating ligand concentration, Equation 31 becomes

$$\Delta\Delta G_{hx} = -kT \log \frac{K_{bd}}{K_{bN}}, \quad (32)$$

which is the result given in Equation 8. If $\Delta\Delta G_{hx}$ is known, Equation 32 can be used to estimate K_{bd}, the binding affinity of the ligand for the particular denatured state, d. This estimate will be less than the *maximum* dL affinity because Equation 32 assumes that *every* exchanging d state binds ligand equally. For monomer 10 of sequence A, only 3% of the exchanging second-excited states binds ligand with the 3-LH contact affinity shown in Figure 8A, so the estimated affinity represents an average that is less than this maximum 3-contact affinity.

HP model of protein flexibility

Here we describe our measure of thermal motion, B in the model proteins. For any monomer, i, of an HP sequence, we can derive a flexibility parameter, B_i, by taking all possible conformations and calculating the average deviation of the monomer from its position in the native state, N. To illustrate, consider a single non-native conformation, X. The "position" of monomer i in this conformation is given by its distance matrix, $d_i(X)$, i.e., the sum of its intra-monomer distances:

$$d_i(X) = \sum_{j \neq i} r_{ij}^2(X), \quad (33)$$

where j is the monomer index and $r_{ij}(X)$ is the distance between monomers i and j, in arbitrary lattice units.

The difference in position of monomer i in conformation X and in the native state, N, is calculated from a relative distance matrix, $B_i(X)$:

$$B_i(X) = \sum_{j \neq i} \frac{|r_{ij}^2(X) - r_{ij}^2(N)|}{(r_{ij}^2(X) + r_{ij}^2(N))/2}, \quad (34)$$

where the denominator of Equation 34 is a normalization factor. The flexibility, B_i, of monomer i is then calculated from the Boltzmann average of $B_i(X)$ over all possible conformations, X.

The quantity B_i represents an *equilibrium* measure of flexibility rather than a *dynamic* one, i.e., flexibility is determined by the populations of low-energy non-native states, not by the heights of kinetic barriers that separate them, which we do not treat here. However, because HX is shown to be an equilibrium process for many proteins, B_i captures the component of flexibility most relevant to HX measurements.